ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

UIT TOÀN DIỆN • SÁNG TẠO • PHỤNG SỰ

BÀI GIẢNG
XÁC SUẤT THỐNG KÊ

Giảng viên: TS. PHÙNG MINH ĐỨC

(Bộ môn Toán Lý)

Chương 2: Biến ngẫu nhiên

2.1 Biến ngẫu nhiên

Định nghĩa 2.1.1

Cho một phép thử có không gian mẫu là Ω. Biến ngẫu nhiên (BNN) X (random variable - đại lượng ngẫu nhiên) một phép gán một kết quả (outcome) của Ω với một số.

Ví dụ 2.1.2

a. Gieo 3 đồng xu và tính số mặt sấp. Đặt X là số mặt sấp. Ta thấy rằng X có thể nhận các giá trị 0,1,2,3.
b. Trong một cái túi có 1 bi đỏ và 99 bi đen. Người chơi cần bỏ ra 10000 đồng để lấy ngẫu nhiên 1 bi. Nếu lấy trúng bi đỏ thì được thưởng 100000 đồng. Nếu lấy trúng bi đen thì không được thưởng. Gọi X là số tiền mà người chơi có sau 1 lần lấy bi. Khi đó X nhận các giá trị 90000 và -10000.

Định nghĩa 2.1.4

1. Biến ngẫu nhiên được gọi là **rời rạc** nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử.
2. Biến ngẫu nhiên được gọi là **liên tục** nếu tập giá trị của nó lấp kín một khoảng trên trục số.

Ví dụ 2.1.5

1. Đặt X là số chấm xuất hiện khi gieo một con xúc xắc. Khi đó X nhận các giá trị là 1,2,3,4,5,6. Do đó, X là một BNN rời rạc.
2. Đặt Y là nhiệt độ trong một ngày. Khi đó Y là một số ngẫu nhiên từ -90^oC đến 71^oC. Do đó Y là một BNN liên tục.

2.2 Biến ngẫu nhiên rời rạc

Định nghĩa 2.2.1

Cho X là một biến ngẫu nhiên rời rạc có miền giá trị là $\{x_1, x_2, \ldots\}$. Bảng phân phối xác suất của biến ngẫu nhiên X là

X	x_1	x_2	$\ldots$
$P(X = x_i)$	p_1	p_2	$\ldots$

trong đó $\sum_{i=1}^{+\infty} p_i = 1$.

Ví dụ 2.2.2

Xét phép thử là gieo một con xúc xắc. Đặt X là "số chấm xuất hiện" và ta thấy nó nhận các giá trị nguyên từ 1 đến 6.
Bảng phân phối của biến ngẫu nhiên X là

X	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Gọi Y là tổng số chấm khi gieo con xúc xắc hai lần. Bảng phân phối của BNN Y là ...

Ví dụ 2.2.3

Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Hãy lập bảng phân phôi xác suất của số đạn cần bắn.

Giải. Gọi A_i là viên đạn thứ i trúng mục tiêu, $i = 1, 2, 3$ và X là số đạn cần bắn. Khi đó $X \in \{1, 2, 3\}$.

- $P(X = 1) = P(A_1) = 0,6.$
- $P(X = 2) = P(\overline{A_1}A_2) = P(\overline{A_1})P(A_2) = 0,4.0,6 = 0,24.$
- $P(X = 3) = 0,4.0,4 = 0,16.$

Bảng phân phối của biến ngẫu nhiên X là

X	1	2	3
$P(X = x_i)$	$0,6$	$0,24$	$0,16$

Ví dụ 2.2.4

Gieo một con xúc xắc và gọi X là số lần gieo đến khi gặp mặt 6 chấm. Lập bảng phân phối xác suất của X.

Giải. Biến ngẫu nhiên $X \in \{1, 2, 3, \ldots\}$ (vô hạn đếm được). Xác suất để lần thứ n xuất hiện mặt 6 chấm là

$$P(X = n) = \left(\frac{5}{6}\right)^{n-1} . \frac{1}{6}.$$

Bảng phân phối xác suất của X là

X	1	2	$\ldots$	n	$\ldots$
$P(X = x_i)$	$\frac{1}{6}$	$\frac{5}{6}.\frac{1}{6}$	$\ldots$	$\left(\frac{5}{6}\right)^{n-1} .\frac{1}{6}$	$\ldots$

Định nghĩa 2.2.5 (Hàm phân phối xác suất)

Hàm phân phối xác suất của biến ngẫu nhiên X, ký hiệu là $F(x)$, được xác định như sau:

$$F(x) = P(X \leq x),\ x \in \mathbb{R}.$$

Cho X là một biến ngẫu nhiên rời rạc nhận các giá trị $\{x_1, x_2, \dots\}$. Nếu $x_1 < x_2 < \dots$ thì

$$F(x_n) = \sum_{i=1}^{n} P(X = x_i).$$

$$F(x) = \sum_{x_i \leq x} P(X = x_i) = \begin{cases} 0 & \text{nếu } x < x_1 \\ p_1 & \text{nếu } x_1 \leq x < x_2 \\ p_1 + p_2 & \text{nếu } x_2 \leq x < x_3 \\ \dots & \dots \end{cases}$$

Ví dụ 2.2.6 Gieo một con xúc xắc. Đặt biến ngẫu nhiên X là "số chấm xuất hiện." Bảng phân phối của biến ngẫu nhiên X là

X	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Hàm phân phối xác suất của X được xác định bởi

$$F(1) = \frac{1}{6}; F(2) = \frac{2}{6}; F(3) = \frac{3}{6}; F(4) = \frac{4}{6}; F(5) = \frac{5}{6}; F(6) = 1$$

Tính $F(0), F(2.2), F(7)$.
Hàm F xác định bởi:

....

....

....

Ví dụ 2.1.8. Có hai lô hàng: lô 1 có 10 sản phẩm trong đó có 3 sản phẩm loại A; lô 2 có 12 sản phẩm trong đó có 4 sản phẩm loại A.

a. Lấy ngẫu nhiên mỗi lô một sản phẩm. Gọi X là số sản phẩm loại A lấy được. Lập bảng phân phối xác suất (luật phân phối xác suất) của X.

b. Lấy ngẫu nhiên một lô hàng, từ đó lấy ngẫu nhiên 3 sản phẩm. Đặt Y là số sản phẩm loại A lấy được. Lập bảng phân phối xác suất của Y.

Hướng dẫn: a. Tập giá trị của X là $\{0, 1, 2\}$.

- $P(X = 0) = \ldots;$
- $P(X = 2) = \ldots;$
- $P(X = 1) = \ldots.$

Bảng phân phối xác suất của X là

X	....	....	....
$P(X = x_i)$	....	....	....

b. Áp dụng công thức xác suất đầy đủ.
Gọi A_i là biến cố "Lấy lô i" với $i = 1, 2$. Ta có

$$P(A_1) = P(A_2) = \dots\dots\dots\dots$$

- $P(Y = 0) = \dots;$
- $P(Y = 1) = \dots;$
- $P(Y = 2) = \dots;$
- $P(Y = 3) = \dots.$

Y	....	....	....	....
$P(Y = y_i)$	....	....	....	....

2.3 Biến ngẫu nhiên liên tục

Định nghĩa 2.3.1 (Hàm mật độ)

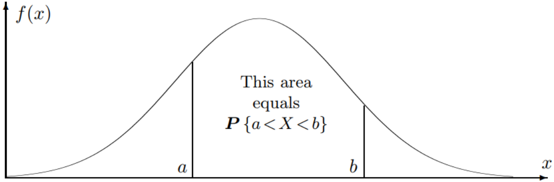
Hàm mật độ xác suất của một biến ngẫu nhiên liên tục là một hàm số $f(x)$ thỏa mãn 2 điều kiện sau:

1. $f(x) \geq 0, \forall x \in (-\infty, +\infty)$
2. $\displaystyle\int_{-\infty}^{+\infty} f(x)dx = 1,$

Nếu $f(x)$ là hàm mật độ của BNN X thì với mọi $a \leq b$, ta có:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

- Hai điều kiện 1 và 2 trong định nghĩa trên có nghĩa là phần diện tích giới hạn bởi đồ thị của $f(x)$ và trục hoành bằng 1.
- Xác suất khi biến ngẫu nhiên X nhận các giá trị thuộc đoạn $[a, b]$ bằng diện tích hình giới hạn bởi của đồ thị $f(x)$, trục hoành và hai đường thẳng $x = a; x = b$.



Định nghĩa 2.3.2 (Hàm phân phối xác suất)

Hàm phân phối xác suất của một biến ngẫu nhiên liên tục X được xác định như sau

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

Định lý 2.3.3

Cho X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và hàm phân phối xác suất $F(x)$. Khi đó

1. $P(X = a) = \int_a^a f(x)dx = 0$
2. $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$.
3. $P(X > a) = 1 - P(X \leq a) = \int_a^{+\infty} f(t)dt.$

Ví dụ 2.3.4

Thời gian sử dụng của một thiết bị điện tử (tính bằng năm) là một biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \dfrac{k}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases}$$

a. Tìm k.
b. Tìm hàm phân phối xác suất.
c. Tìm xác suất thiết bị đó sử dụng được hơn 5 năm.

a. Tìm k. Ta có

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{1} f(x)dx + \int_{1}^{+\infty} f(x)dx = \ldots \Rightarrow k = \ldots$$

b. Hàm phân phối xác suất

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0, \text{ với } x < 1 \\ \int_1^x \frac{k}{t^3}dt = \ldots, \text{ với } x \geq 1 \end{cases}$$

c. Xác suất để thiết bị đó sử dụng được nhiều hơn 5 năm là

$$P(X > 5) = \int_{5}^{+\infty} f(x)dx = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

(Ta có thể sử dụng $P(X > 5) = 1 - F(5) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$)

Bài tập 2.3.5

Thời gian phục vụ mỗi khách hàng tại một cửa hàng là một biến ngẫu nhiên liên tục X (tính bằng phút) có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} k.e^{-4x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

a. Tìm k và hàm phân phối xác suất của X.
b. Tính xác suất thời gian phục vụ một khách hàng trong khoảng thời gian từ 1 đến 2 phút.

Ký hiệu: $\exp(-4x) = e^{-4x}$.

Bài tập 2.3.6

Cho biến ngẫu nhiên X là thời gian sử dụng của một loại pin (tính bằng năm). Giả sử hàm mật độ xác suất của X được cho bởi

$$f(x) = \begin{cases} \frac{1}{2}\exp(-\frac{x}{2}), & x > 0 \\ 0, & x \leq 0. \end{cases}$$

a. Tính xác suất để loại pin này sử dụng được ít hơn 2 năm hoặc nhiều hơn 4 năm.
b. Giả sử có một viên pin loại này đã được sử dụng 2 năm. Tính xác suất để pin này được sử dụng ít nhất 3 năm.

Nếu biết hàm phân phối xác suất $F(x)$ thì ta tìm hàm được mật độ xác suất

$$f(x) = F'(x).$$

Ví dụ 2.3.7 Đặt X là thời gian (tháng) sử dụng của một loại thiết bị điện tử. Giả sử biến ngẫu nhiên X có hàm phân phối xác suất

$$F(x) = \begin{cases} 1 - e^{-0{,}03x^{1{,}2}}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

a. Chọn ngẫu nhiên một thiết bị. Tính xác suất thiết bị đó sử dụng được ít nhất 12 tháng.
b. Tìm hàm mật độ xác suất của X.

2.4 Các tham số của biến ngẫu nhiên

Định nghĩa 2.4.1 (Kỳ vọng)

Kỳ vọng (expectation) của biến ngẫu nhiên X, ký hiệu là $E(X)$, được xác định như sau:

- Nếu X là biến ngẫu nhiên rời rạc thì

$$E(X) = \sum_{i \geq 1} x_i P(X = x_i)$$

- Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Ta thường ký hiệu $E(X) = \mu$.

Kỳ vọng còn được gọi là giá trị trung bình (Mean), giá trị trung bình có trọng số (Weighted Average), giá mong đợi (Expected Value).

Ý nghĩa. Kỳ vọng của biến ngẫu nhiên là **giá trị trung bình có trọng số** của biến ngẫu nhiên (phân biệt với trung bình cộng của các giá trị).

Ví dụ 2.4.2

Trong một hộp có 99 tấm thẻ đen và 1 tấm thẻ đỏ. Mỗi lần rút thẻ người chơi sẽ phải trả 10000 đồng. Nếu rút trúng thẻ màu đỏ thì người chơi được thưởng 100000 đồng.

a. Nếu người chơi chỉ rút 1 lần thì kỳ vọng số tiền thu được bằng bao nhiêu?

b. Nếu người chơi rút 2 lần (rút lần thứ nhất rồi bỏ lại vào hộp) thì kỳ vọng số tiền thu được bằng bao nhiêu?

a. Gọi X là "Số tiền thu được sau khi rút". Nếu rút trúng thẻ đỏ thì $X = 90000$. Nếu rút được thẻ đen thì $X = -10000$. Bảng phân phối xác suất

X	90000	-10000
$P(X = x_i)$	$\frac{1}{100}$	$\frac{99}{100}$

Kỳ vọng $E(X) = 90000.\dfrac{1}{100} + (-10000).\dfrac{99}{100} = -9000$

(Trung bình mỗi lần chơi, người chơi mất 9000 đồng).

b. Gọi Y là "Số tiền thu được sau hai lần rút". Khi đó Y nhận các giá trị: $180000, 80000, -20000$. Bảng phân phối xác suất của Y là

Y	180000	80000	-20000
$P(Y = y_i)$	$\frac{1}{10000}$	$\frac{198}{10000}$	$\frac{9801}{10000}$

Kỳ vọng của Y là

$$E(Y) = 180000.\frac{1}{10000} + 80000.\frac{198}{10000} + (-20000).\frac{9801}{10000} = -18000.$$

(Trung bình sau 2 lần chơi, người chơi sẽ mất 18000 đồng)

Ví dụ 2.4.3

Gọi X là lượng cát (tính bằng nghìn tấn) do công ty CAT bán ra hàng tuần. Giả sử biến ngẫu nhiên X có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} k(1-x^2), 0 \leq x \leq 1 \\ 0, x \notin [0;1] \end{cases}$$

a. Tìm k.

b. Tính trung bình lượng cát bán ra hàng tuần của công ty đó.

Giải:a.

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} + \int_{0}^{1} + \int_{1}^{+\infty} = \ldots \Rightarrow k = 3/2$$

b. Kỳ vọng của X (trung bình lượng cát bán ra hàng tuần) là

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{1} \frac{3}{2}x(1-x^2)dx = \frac{3}{8}$$

Định nghĩa 2.4.4 (Phương sai)

Cho X là biến ngẫu nhiên và kỳ vọng μ. Phương sai (variance) của X, kí hiệu là $V(X)$ (hoặc DX, σ^2,) được xác định bởi

1. Nếu X là biến ngẫu nhiên rời rạc có miền giá trị $\{x_1, x_2, \ldots\}$ thì

$$V(X) = \sigma^2 = \sum_{i=1}^{\infty}(x_i - \mu)^2 P(X = x_i) = E((X-\mu)^2)$$

2. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì

$$V(X) = \sigma^2 = \int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx = E((X-\mu)^2)$$

Độ lệch chuẩn (standard deviation) của X được ký hiệu và xác định bởi Std$(X) = \sqrt{V(X)} = \sigma$.

Định lý 2.4.5

1. Nếu X là một biến ngẫu nhiên rời rạc và $g(x)$ là một hàm số thì
$$E(g(X)) = \sum_x g(x)P(X = x)$$
2. Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và $g(x)$ là một hàm số thì
$$E(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx.$$

Mệnh đề 2.4.6

Cho X là một biến ngẫu nhiên và $a, b \in \mathbb{R}$. Khi đó

1. $Eb = b$ và $E(aX + b) = aE(X) + b$
2. $V(b) = 0$ và $V(aX + b) = a^2 V(X)$
3. $V(X) = E(X^2) - (EX)^2$

Ví dụ 2.4.7

Một trung tâm bảo hành laptop thấy rằng số laptop hiệu Potpal cần sửa trong tuần tối đa là 3 máy. Cho X là số latop Potpal cần sửa trong một tuần. Dựa trên các số liệu đã có, bảng phân phối xác suất của X như sau

X	0	1	2	3
$P(X = x_i)$	0,2	0,3	0,4	0,1

a. Trong một tuần, trung bình có bao nhiêu laptop Potpal cần sửa?
b. Tìm phương sai, độ lệch chuẩn?.
c. Một cửa hàng Z có hợp đồng với trung tâm bảo hành đối với việc sửa laptop Potpal như sau: Phí dịch vụ hàng tuần là 20$ và 5$ cho mỗi laptop được sửa. Tìm trung bình, phương sai và độ lệch chuẩn của số tiền mà cửa hàng Z trả cho trung tâm bảo hành mỗi tuần.

a. Kỳ vọng của X là

$$\mu = E(X) = \sum_{i=1}^{4} x_i P(X = x_i) = 0.0,2 + 1.0,3 + 2.0,4 + 3.0,1 = 1,4.$$

Như vậy, số laptop cần sửa trung bình là 1,4 máy trong một tuần.
b. Phương sai của X là

$$V(X) = \sum_{i=1}^{4} (x_i - \mu)^2 P(X = x_i) = \cdots = 0,84$$

(hoặc tính $V(X) = E(X^2) - \mu^2 = \cdots = 0,84$) và độ lệnh chuẩn là

$$\sigma = \sqrt{V(X)} = \sqrt{0,84} = 0,92.$$

c. Chi phí hàng tuần mà của hàng Z phải trả cho trung tâm bảo hành là một hàm số theo biến X như sau

$$20 + 5X.$$

Do đó

$$E(20 + 5X) = 5EX + 20 = 5.1,4 + 20 = 27$$

Như vậy, cửa hàng Z trả cho trung tâm bảo hành trung bình 27$ mỗi tuần.
Phương sai của chi phí hàng tuần là

$$V(20 + 5X) = 5^2V(X) = 5^2.0,84 = 21$$

và độ lệnh chuẩn là

$$\sigma = \sqrt{21} = 4,583.$$

Ví dụ 2.4.8 Một cửa hàng bán một loại hóa chất thấy rằng lượng hóa chất bán ra hàng tuần (đơn vị tính là trăm lít) là X có hàm mật độ xác suất là

$$f(x) = \begin{cases} \frac{x}{4}, & 0 \leq x \leq 2 \\ \frac{1}{2}, & 2 < x \leq 3 \\ 0, & \text{các trường hợp khác} \end{cases}$$

Giả sử mỗi tuần, chi phí bảo dưỡng cửa hàng là 50$. Cửa hàng mua mỗi lít hóa chất là 1,75$ và bán ra 2,7$.

a. Tính lượng hóa chất trung bình bán ra hàng tuần và độ lệch chuẩn tương ứng?

b. Tính lợi nhuận trung bình hàng tuần của cửa hàng và độ lệch chuẩn tương ứng?

....

....

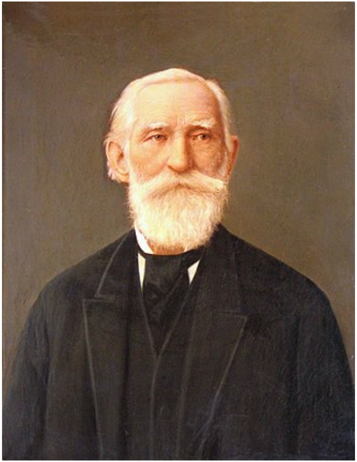
Theorem 1 (Định lý Chebyshev)

Cho X là một biến ngẫu nhiên có kỳ vọng μ và phương sai σ^2. Khi đó, với mọi $\varepsilon > 0$, ta có

$$P(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Định lý Chebyshev cho thấy

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$



Pafnuty Lvovich Chebyshev (1821 - 1894) là nhà toán học người Nga.

Ví dụ 2.4.9

Khảo sát tình hình sản xuất trong năm 2021, giám đốc một nhà máy thấy rằng trung bình một ngày làm được 120 sản phẩm với độ lệch chuẩn là 10.
a. Tỉ lệ những ngày nhà máy sản xuất được từ 100 đến 140 sản phẩm có nhiều hơn 70% không?
b. Tìm khoảng số lượng sản phẩm ngắn nhất để chứa ít nhất 90% mức sản xuất hàng ngày.

Giải.
a. Đặt X là số sản phẩm sản xuất trong một ngày. Ta có

$$E(X) = \mu = 120, \sigma = \sqrt{V(X)} = 10.$$

Vì $X \in [100; 140] = [\mu - \varepsilon, \mu + \varepsilon]$ nên $\varepsilon = 20$.

Khi đó, theo Định lý Chebyshev,

$$P(100 \leq X \leq 140) \geq 1 - \frac{10^2}{20^2} = \frac{3}{4} = 75\%$$

Như vậy, tỉ lệ những ngày làm được từ 100 đến 140 sản phẩm không nhỏ hơn 75%. Do đó, tỉ lệ này lớn hơn 70%.
b. Ta cần tìm ε sao cho

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 90\%$$

hay

$$1 - \frac{\sigma^2}{\varepsilon^2} = 0,9 \Rightarrow \varepsilon = \sqrt{1000} \approx 31,6.$$

Suy ra

$$X \in [120 - 31,6; 120 + 31,6] = [88,4; 151,6].$$

Định nghĩa 2.4.10 (Mode)

Mode của biến ngẫu nhiên X là một giá trị của biến ngẫu nhiên X, ký hiệu Mod(X), mà tại đó biến ngẫu nhiên X nhận xác suất lớn nhất (nếu X là rời rạc) hoặc tại đó hàm mật độ đạt giá trị lớn nhất (nếu X là biến liên tục).

Như vậy, $Mod(X) = x^*$ là giá trị xác định bởi:

- $p^* = \max\{p_i, i = 1, 2, \ldots\}$ nếu X rời rạc.
- $f(x^*) = \max\{f(x), x \in (-\infty, +\infty)\}$ nếu X liên tục và có hàm mật độ $f(x)$.

Định nghĩa 2.4.11 (Trung vị)

Trung vị (median) là giá trị của biến ngẫu nhiên X mà nó chia phân phối của biến ngẫu nhiên thành hai phần bằng nhau, kí hiệu Med(X).

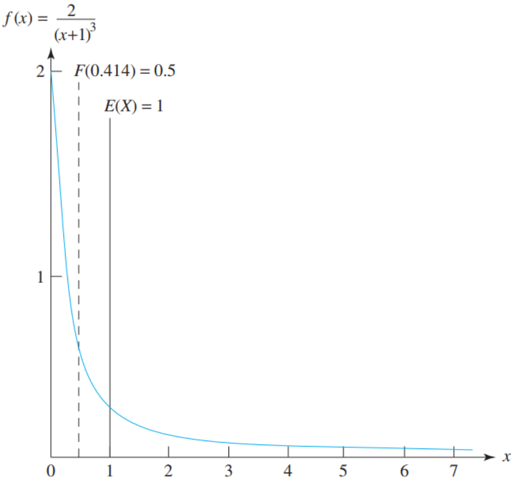
Nếu X là biến ngẫu nhiên rời rạc thì giá trị x_k sẽ là trung vị nếu điều kiện sau được thỏa mãn:

$$F(x_{k-1}) < 0,5 \leq F(x_k).$$

Nếu X là biến ngẫu nhiên liên tục thì trung vị của X là giá trị m thỏa mãn điều kiện:

$$\int_{-\infty}^{m} f(x)dx = 0,5.$$

Hình minh họa trung vị và kỳ vọng



2.5 Các phân phối xác suất thường gặp

Định nghĩa 2.5.1 (Phép thử Bernoulli)

Một **phép thử Bernoulli** là phép thử mà ta chỉ quan tâm đến hai biến cố: A (thành công) và $\overline{A}$ (thất bại) trong đó $P(A) = p$.

Định nghĩa 2.5.2 (Phân phối Bernoulli)

Một phân phối xác suất của biến ngẫu nhiên rời rạc X được gọi là phân phối Bernoulli nếu X chỉ nhận hai giá trị 1 (nếu thành công) hoặc 0 (nếu thất bại), trong đó $P(X = 1) = p$ (gọi là xác suất thành công) và $P(X = 0) = 1 - p$ (gọi là xác suất thất bại). Ký hiệu $X \sim B(p)$.

Bảng phân phối xác suất của biến ngẫu nhiên X có phân phối Bernoulli là

X	0	1
$P(X = x_i)$	$1-p$	p

Cho biến ngẫu nhiên $X \sim B(p)$. Khi đó

$$E(X) = p \text{ và } V(X) = (1-p)p.$$

Ví dụ 2.5.3

Một câu hỏi trắc nghiệm có 4 phương án trả lời trong đó có một phương án đúng. Một sinh viên chọn ngẫu nhiên một phương án để trả lời câu hỏi đó. Đặt A là biến cố "sinh viên chọn phương án đúng". Khi đó việc trả lời của sinh viên là một phép thử Bernoulli và $p = P(A) = \frac{1}{4}$ và $P(\overline{A}) = \frac{3}{4}$.

Gọi X là biến ngẫu nhiên xác định bởi

$$X = \begin{cases} 1 \text{ nếu sinh viên trả lời đúng} \\ 0 \text{ nếu sinh viên trả lời sai.} \end{cases}$$

Khi đó $X \sim B(\frac{1}{4})$ và $E(X) = \frac{1}{4}, V(X) = \frac{3}{4}.\frac{1}{4} = \frac{3}{16}$.

Định nghĩa 2.5.4 (Dãy phép thử Bernoulli)

Dãy các phép thử lặp lại, độc lập, trong mỗi phép thử chỉ có một trong hai biến cố A và $\overline{A}$, xác suất $P(A) = p$ không đổi được gọi là dãy **phép thử Bernoulli.**

Xác suất biến cố A xảy ra k lần trong n phép thử là:

$$\boxed{P_n(k) = C_n^k p^k (1-p)^{n-k}}$$

Công thức nêu trên được gọi là **công thức Bernoulli.**

Định nghĩa 2.5.7 (Phân phối nhị thức)

Xét dãy n phép thử Bernoulli độc lập và xác suất của biến cố A trong mỗi phép thử là p. Đặt X là số lần xuất hiện biến cố A trong n phép thử. Ta nói X là một biến ngẫu nhiên có **phân phối nhị thức**, ký hiệu $X \sim B(n,p)$, với các tham số p, n. Xác suất trong n phép thử có k lần xuất hiện A là

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, 2, \ldots, n.$$

Mệnh đề 2.5.8

Cho biến ngẫu nhiên X có phân phối nhị thức $X \sim B(n,p)$. Khi đó

$$E(X) = np \text{ và } V(X) = n(1-p)p.$$

Trong dãy phép thử Bernoulli, số $x_0 = Mod(X)$ gọi là **số có khả năng nhất.**

- Nếu $(n+1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = (n+1)p - 1$ và $x_0 = (n+1)p$.
- Nếu $(n+1)p \notin \mathbb{Z}$ thì $x_0 = [(n+1)p - 1] + 1$, trong đó $[(n+1)p - 1]$ là phần nguyên của $(n+1)p - 1$.

Ví dụ 2.5.6 Tỷ lệ mắc Covid-19 ở một thành phố nọ là 10%. Trong đợt khám bệnh cho vùng đó, người ta đã khám 100 người. Trung bình trong 100 người được khám có bao nhiêu người nhiễm bệnh?Tìm số người bị Covid-19 có khả năng nhấ và tính xác suất tương ứng.

Giải. Gọi X là số người nhiễm bệnh trong 100 người được khám. Khi đó $X \sim B(100, 0.1)$.

Trong 100 người được khám, số người nhiễm bệnh trung bình là:

$$EX = 100 \times 0.1 = 10.$$

Ta có $(n+1)p = (100+1).0,1 \notin \mathbb{Z}$. Do đó, số có khả năng lớn nhất là

$$x_0 = [(n+1)p - 1] + 1 = [9,1] + 1 = 9 + 1 = 10$$

và xác suất tương ứng là $P(X = 10) = C_{100}^{10}.0,1^{10}.0,9^{90} \approx 0,1319.$

Ví dụ 2.5.9

Một đề thi môn xác suất thông kê có 20 câu trắc nghiệm, đáp án mỗi câu gồm 4 phương án trong đó có 1 phương án đúng. Sinh viên B chọn ngẫu nhiên các phương án trong mỗi câu trả lời. Nếu trả lời đúng 1 câu thì sinh viên B được 0,5 điểm và nếu trả lời sai 1 câu thì được 0 điểm. Tính xác suất sinh viên được đúng 5,5 điểm.

Giải. Gọi X là số câu đúng trong 20 câu. Khi đó $X \sim B(20, 0.25)$. Để được đúng 5,5 điểm, sinh viên B phải chọn đúng 11 câu trả lời đúng và 9 câu trả lời sai. Xác suất B được 5,5 điểm là

$$P(X = 11) = C_{20}^{11} \left(\frac{1}{4}\right)^{11} \left(\frac{3}{4}\right)^{9} = 0,003.$$

Khả năng cao nhất là SV đó được mấy điểm?

Định nghĩa 2.5.10 (Phân phối siêu bội)

Xét một tập có N phần tử trong đó có M phần tử có tính chất $\mathcal{P}$ và $N - M$ phần tử không có tính chất $\mathcal{P}$. Chọn ra n phần tử từ N phần tử. Gọi X là số phần tử có tính chất $\mathcal{P}$ trong n phần tử đã chọn. Khi đó X là một biến ngẫu nhiên và ta nói X có **phân phối siêu bội** với ba tham số N, M, n. Ký hiệu $X \sim H(N, M, n)$

Xác suất trong n phần tử có k phần tử có tính chất $\mathcal{P}$ là

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

trong đó $\max\{0, M + n - N\} \leq k \leq \min\{n, M\}$.

Ví dụ 2.5.11 Công ty Z đã nhận được 20 mẫu chào hàng cung cấp với máy in, trong đó 8 máy in PH và 12 máy in Nonac. Chọn ngẫu nhiên 5 mẫu chào hàng. Tính xác suất trong 5 mẫu chào hàng có k ($k = 0, 1, 2, 3, 4$, hoặc 5) máy in Nonac?

Giải. Ta có một tập gồm $N = 20$ phần tử, trong đó có $M = 12$ máy in Nonac và $N - M = 8$ máy in PH. Chọn ra 5 mẫu chào hàng, tức là $n = 5$. Đặt X là số máy in Nonac trong 5 mẫu chào hàng. Khi đó $X \sim H(20, 12, 5)$ và

$$P(X = k) = \frac{C_{12}^{k} C_{8}^{5-k}}{C_{20}^{5}}.$$

Mệnh đề 2.5.12

Kỳ vọng và phương sai của một phân phối siêu bội $X \sim H(N, M, n)$ được xác định như sau:

$$E(X) = n.\frac{M}{N} \text{ và } V(X) = \frac{N-n}{N-1}.n.\frac{M}{N}.\left(1 - \frac{M}{N}\right)$$

Ví dụ 2.5.13 Trong 100 laptop trưng bày tại cửa hàng POTPAL có 70 laptop hiệu PH. Chọn ngẫu nhiên 40 laptop trong số 100 laptop. Gọi X là số laptop hiệu PH trong số 40 laptop được chọn.
a. Tính xác suất để chọn được từ 27 đến 29 laptop PH.
b. Tính trung bình số laptop PH được chọn và $V(X)$.

....

....

....

Định nghĩa 2.5.14 (Phân phối Poisson)

Một biến ngẫu nhiên rời rạc X được gọi là có **phân phối Poisson** với tham số $\lambda(\lambda > 0)$, ký hiệu $X \sim P(\lambda)$, nếu nó có phân phối xác suất là

$$P(X = x) = \frac{e^{-\lambda}.\lambda^x}{x!}, x = 0, 1, 2, \dots$$

trong đó λ là trung bình số lần xuất hiện biến cố và x là số lần xuất hiện biến cố mà ta quan tâm.

Mệnh đề 2.5.16

Cho X là một biến ngẫu nhiên có phân phối Poisson $X \sim P(\lambda)$. Khi đó

$$E(X) = V(X) = \lambda.$$

Giá trị của hàm phân phối xác suất $F(X) = P(X \leq x)$ của biến ngẫu nhiên có phân phối Poisson có thể tìm trong bảng.
Phân phối Poisson là một phân phối xác suất của biến ngẫu nhiên rời rạc. Thông tin mà ta biết **không phải** là **xác suất của một biến cố** trong một lần thử như trong phân phối Bernoulli hay là **số lần xảy ra biến cố** đó trong n lần thử như trong phân phối nhị thức, mà **ta biết** *trung bình số lần xảy ra của một biến cố trong một khoảng thời gian nhất định.* Giá trị trung bình này được kí hiệu là μ. Phân phối Poisson còn được dùng cho khoảng mà đơn vị khác thời gian như: khoảng cách, diện tích hay thể tích.

Siméon Denis Poisson (1781 - 1840) là nhà toán học, nhà vật lý người Pháp. Ảnh nguồn: Internet.

Ví dụ 2.5.17 Giả sử rằng số lần truy cập trung bình vào một trang web là 2 lần trong 1 giây.

a. Tính xác suất có 5 lần truy cập trong 1 giây.

b. Tính xác suất có nhiều nhất 5 lần truy cập trong 1 giây.

c. Tính xác suất có nhiều hơn 5 lần truy cập trong 2 giây.

Giải. a. Đặt X là số lần truy cập trong 1 giây. Khi đó X có phân phối Poisson với tham số $\lambda = 2$. Suy ra

$$P(X = 5) = \frac{e^{-2}.2^5}{5!} = 0,0361.$$

b. Tra bảng, ta có $P(X \leq 5) = 0,983$.

c. Trung bình số lần truy cập trong 2 giây là 4. Đặt Y là số lần truy cập vào trang web trong 2 giây. Khi đó $Y \sim P(4)$ và do đó

$$P(Y > 5) = 1 - P(Y \leq 5) = 1 - 0,785 = 0,215$$

Ví dụ 2.5.18 Một bệnh viện có trung bình 10 ca sinh mỗi giờ. Đặt X là số ca sinh trong một giờ. Khi đó X là một biến ngẫu nhiên có phân phối Poisson. Tính

a. Xác suất để có nhiều hơn 12 lần sinh xảy ra trong một giờ nhất định là bao nhiêu?

b. Xác suất để xảy ra ít hơn 5 ca sinh trong một giờ nhất định?

c. Xác suất để từ 8 đến 11 ca sinh xảy ra trong một giờ nhất định là bao nhiêu?

Ví dụ 2.5.19 Số lượng cơn bão lớn trung bình trong thành phố T là 2 cơn bão mỗi năm. Xác suất chính xác 3 cơn bão sẽ đổ bộ vào thành phố T trong năm tới là bao nhiêu?

- Cho $a < b$. BNN X có hàm mật độ

$$f(x) = \begin{cases} 0 & \text{nếu } x < a \text{ hoặc } x > b \\ \dfrac{1}{b-a} & \text{nếu } a \leq x \leq b \end{cases}$$

gọi là BNN có **phân phối đều** trên $[a, b]$.

- BNN có phân phối đều trên $[a, b]$ có hàm phân phối

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{nếu } x \leq a \\ \dfrac{x-a}{b-a} & \text{nếu } a < x \leq b \\ 1 & \text{nếu } x > b \end{cases}$$

và

$$EX = \frac{a+b}{2}, \; DX = \frac{(b-a)^2}{12}$$

Định nghĩa 2.5.19 (Phân phối chuẩn-Normal distribution)

Một biến ngẫu nhiên liên tục X được gọi là có **phân phối chuẩn** nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

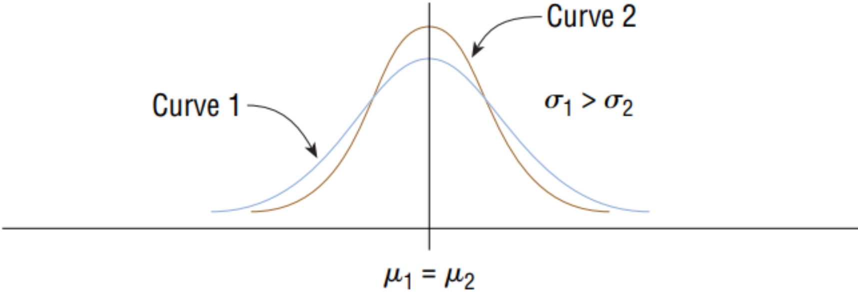
Ký hiệu $X \sim N(\mu; \sigma^2)$.

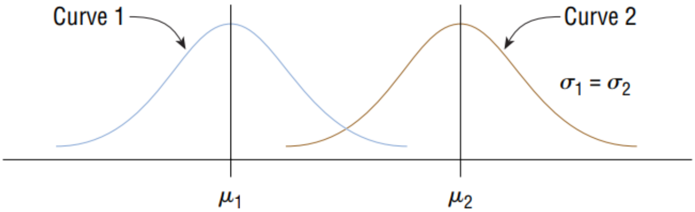
Mệnh đề 2.5.20

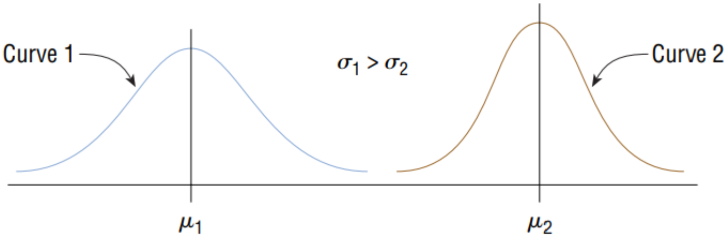
Cho $X \sim N(\mu; \sigma^2)$. Khi đó

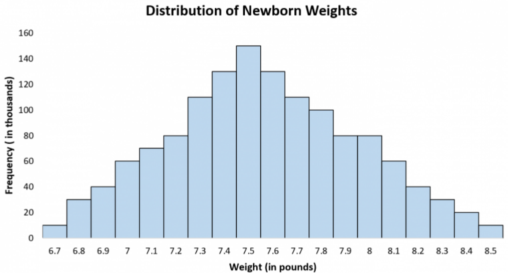
$$E(X) = \mu \text{ và } V(X) = \sigma^2.$$

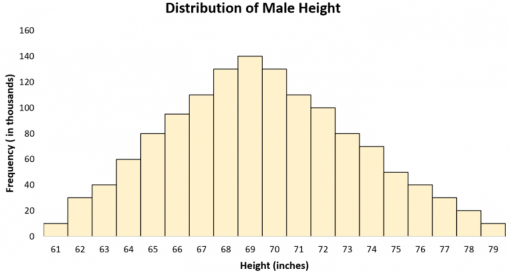
Hình dạng và vị trí của đường cong phân phối chuẩn (đồ thị của hàm mật độ xác suất) phụ thuộc vào hai tham số μ và σ.




Curve 1
Curve 2
$\sigma_1 = \sigma_2$
μ_1
μ_2

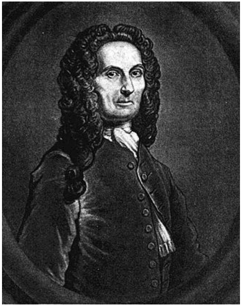

Curve 1
$\sigma_1 > \sigma_2$
Curve 2
μ_1
μ_2





Nguồn: https://www.statology.org

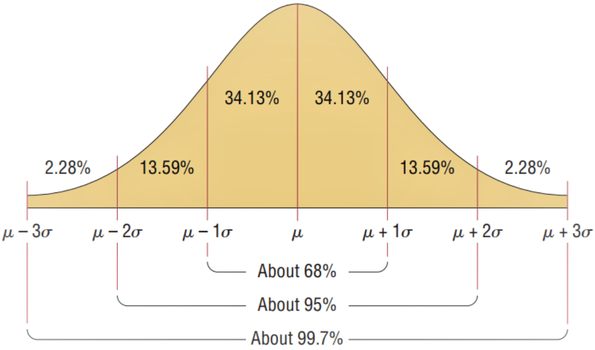
Phân phối chuẩn được nêu ra bởi một người Anh gốc Pháp tên là Abraham de Moivre (1733). Sau đó, Gauss, một nhà toán học người Đức, đã dùng phân phối chuẩn để nghiên cứu các dữ liệu về thiên văn học (1809) và do vậy nó cũng được gọi là phân phối Gauss.

Johann Carl Friedrich Gauss (1777-1855) và Abraham de Moivre (1667-1754). Nguồn: Wikipeadia.org

Một số tính chất của phân phối chuẩn

1. Kỳ vọng, trung vị bằng nhau và nằm ở trung điểm của phân phối.
2. Đồ thị của hàm mật độ xác suất có hình chuông, đối xứng qua đường thẳng đứng đi qua kỳ vọng.
3. Đường cong liên tục và không chạm vào trục x.
4. Diện tích dưới phần của đường cong phân phối chuẩn trong đoạn $[\mu - \sigma; \mu + \sigma]$ là khoảng 0,68; trong đoạn $[\mu - 2\sigma; \mu + 2\sigma]$ là 0,95; và trong khoảng $[\mu - 3\sigma; \mu + 3\sigma]$ là khoảng 0,997.


34.13%
34.13%
2.28%
13.59%
13.59%
2.28%
μ − 3σ
μ − 2σ
μ − 1σ
μ
μ + 1σ
μ + 2σ
μ + 3σ
About 68%
About 95%
About 99.7%

Định nghĩa 2.5.21

Phân phối chuẩn $N(0;1)$ được gọi là **phân phối chuẩn tắc** (standard normal distribution).

Phân phối chuẩn $X \sim N(\mu;\sigma^2)$ có thể được chuyển thành phân phối chuẩn tắc bằng cách đổi biến

$$Z = \frac{X-\mu}{\sigma} \sim N(0;1).$$

Khi đó biến ngẫu nhiên Z có hàm mật độ xác suất $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ và hàm phân phối xác suất

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

1. Đổi biến $Z = \dfrac{X-\mu}{\sigma} \Rightarrow Z \sim N(0;1)$.
2. $P(X \leq a) = P(Z \leq \dfrac{a-\mu}{\sigma}) = \Phi(\dfrac{a-\mu}{\sigma})$

Một số công thức cho $N(0,1)$ và $N(\mu, \sigma^2)$:

$Z \sim N(0,1)$	$X \sim N(\mu, \sigma^2)$
$EZ = 0$	$EX = \mu$
$DZ = 1$	$DX = \sigma^2$
$P(Z < a) = \Phi(a)$	$P(X < a) = \Phi(\dfrac{a-\mu}{\sigma})$
$P(Z > a) = 1 - \Phi(a)$	$P(X > a) = 1 - \Phi(\dfrac{a-\mu}{\sigma})$
$P(a < Z < b) = \Phi(b) - \Phi(a)$	$P(a < X < b) = \Phi(\dfrac{b-\mu}{\sigma}) - \Phi(\dfrac{a-\mu}{\sigma})$

Ví dụ 2.5.22 Cho BNN X có phân phối chuẩn tắc $X \sim N(0;1)$:
a. $P(X \leq 1,35) = \Phi(1,35) = 0,9115$.
b. $P(-1,37 \leq X \leq 1,68) = \Phi(1,68) - \Phi(-1,37) = 0,9535 - 0,0853 = 0,8682$.

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900$ và độ lệch chuẩn là 200$. Gặp ngẫu nhiên một gia đình.
a. Tính xác suất gia đình này có thu nhập từ 600$ đến 1200$.
b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất. Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Gọi X là thu nhập của một gia đình.
Theo giả thiết ta có $X \sim N(900; 200^2)$ với $\mu = 900$ và $\sigma = 200$.
Đặt $Z = \dfrac{X - 900}{200} \sim N(0,1)$, khi đó

$$\begin{aligned}
P(600 \leq X \leq 1200) &= P\left(\frac{600-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{1200-\mu}{\sigma}\right) \\
&= P(-1,5 \leq Z \leq 1,5) \\
&= \Phi(1,5) - \Phi(-1,5) \\
&= 0,9332 - 0,0668 = 0,8664.
\end{aligned}$$

Xác suất gia đình này có thu nhập từ 600\$ đến 1200\$ là 0,8664.
Ta cũng có thể nói tỉ lệ các gia đình có thu nhập từ 600\$ đến 1200\$ là 86,64%.

b. Gọi t là thu nhập (tính bằng $)) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \leq t) = P(Z \leq \frac{t-\mu}{\sigma}) = \Phi(\frac{t-\mu}{\sigma}) = 0,003.$$

Tra bảng, ta được $0,003 = \Phi(-2,75)$. Như vậy

$$P(X \leq t) = P(Z \leq \frac{t-\mu}{\sigma}) = \Phi(\frac{t-\mu}{\sigma}) = 0,003 = \Phi(-2,75)$$

Suy ra

$$\frac{t-900}{200} = -2,75 \Rightarrow t = 350.$$

Như vậy, các gia đình này có thu nhập không quá 350$.

Bài tập 1 Một thành phố lắp đặt đèn điện để chiếu sáng đường. Loại đèn này có tuổi thọ trung bình là 1000 giờ với độ lệch chuẩn là 200 giờ. Giả sử tuổi thọ của loại đèn này có phân phối chuẩn.

a. Tính xác suất đèn này bị hỏng trong 700 giờ.

b. Tính xác suất đèn sử dụng được từ 900 giờ đến 1300 giờ.

c. Sau bao lâu thì còn 10% đèn sử dụng được?

Bài tập 2 Các nghiên cứu cho thấy quãng đường các loại xe ôtô cỡ nhỏ chạy được với 1 lít xăng có phân phối chuẩn với trung bình là 15,1 km/lít và độ lệch chuẩn là 1,92 km/lít.

a. Có bao nhiêu phần trăm xe cỡ nhỏ chạy được nhiều hơn 17 km/lít?

b. Các xe tiết kiệm xăng nhất chiếm 7,15%. Quãng đường mà những xe này chạy với 1 lít xăng ít nhất là bao nhiêu?

Định nghĩa 2.5.24 (Hàm gamma)

Cho $\alpha > 0$, hàm gamma $\Gamma(\alpha)$ được xác định bởi

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \mathrm{e}^{-x} dx$$

Mệnh đề 2.5.25

1. Với $\alpha > 1$, ta có $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
2. Với n là một số nguyên dương, ta có $\Gamma(n) = (n-1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Định nghĩa 2.5.26 (Phân phối χ^2)

Cho ν là một số nguyên dương. Một biến ngẫu nhiên X được gọi là có phân phối χ^2 (chi bình phương) với tham số ν, ký hiệu $X \sim \chi^2_\nu$, nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \begin{cases} \dfrac{x^{\frac{\nu}{2}-1}e^{\frac{-x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Tham số ν được gọi là **bậc tự do** của X.

Nếu các biến ngẫu nhiên $X_i \sim N(0;1)$ $(1 \leq i \leq n)$ độc lập thì biến ngẫu nhiên

$$Y = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2_n.$$

Định nghĩa 2.5.27 (Phân phối Student)

Một biến ngẫu nhiên liên tục X được gọi là có phân phối Student với bậc tự do $\nu - 1$, ký hiệu $X \sim St(\nu)$, nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Ta có thể viết $X \sim \dfrac{N(0;1)}{\sqrt{\chi^2_\nu/\nu}}$.

Phân phối Student sẽ được dùng trong việc ước lượng khoảng của trung bình trong chương 5.

William Sealey Gosset. Nguồn: wikipedia.org

William Sealey Gosset (1876-1937) học toán và hóa học tại Đại học Oxford ở Anh. Năm 1899, ông chuyển đến Dublin, Ireland và làm nhân viên thống kê cho nhà máy bia Guinness. Trong quá trình nghiên cứu chất lượng lúa mạch và hoa bia, Gosset đề xuất việc sử dụng t-phân phối. Năm 1908, ông đã công bố ý tưởng của mình trong một bài báo khoa học bằng bút danh "Student" vì Guinness cấm nhân viên của mình xuất bản kết quả nghiên cứu của riêng họ. Do đó, t-phân phối thường được gọi là phân phối Student.

Định nghĩa 2.5.28 (Phân phối Fisher)

Cho hai biến ngẫu nhiên độc lập có phân phối χ^2 với các bậc tự do lần lượt là ν_1 và ν_2. Khi đó tỉ số

$$\frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$$

được gọi là F-phân phối hay phân phối Fisher (F-distribution), ký hiệu là F_{ν_1,ν_2}.

Ronald Aylmer Fisher năm 1913. Nguồn: wikipedia.org

Ronald Aylmer Fisher (1890 - 1962) là một nhà một nhà toán học, thống kê, nhà di truyền học và học thuật người Anh.

2.6 Hàm đặc trưng

Định nghĩa 2.6.1

Hàm đặc trưng của biến ngẫu nhiên X là một hàm số $\phi(t) : \mathbb{R} \to \mathbb{C}$ xác định bởi

- Nếu X là biến ngẫu nhiên rời rạc thì

$$\phi(t) = \sum_{j \geq 1} e^{itx_j} P(X = x_j)$$

- Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) \mathrm{d}x,$$

trong đó $e^{itx} = \cos tx + i \sin tx$ và $i^2 = -1$.

Phân phối của một biến ngẫu nhiên có thể được xác định hoàn toàn thông qua hàm đặc trưng của nó. Hàm đặc trưng cũng cho ta một tiêu chuẩn để xác định xem khi nào biến ngẫu nhiên có hàm mật độ.

Biến ngẫu nhiên có phân phối	Tham số	Hàm đặc trưng
Bernoulli	p	$\phi(t) = 1 - p + pe^{it}$
Nhị thức	n, p	$\phi(t) = (1 - p + pe^{it})^n$
Siêu bội	p	$\phi(t) = \dfrac{p}{1 - e^{it}(1-p)}$
Poisson	μ	$\phi(t) = e^{\mu(e^{it}-1)}$
Chuẩn	μ, σ^2	$\phi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$
Chi-bình phương	ν	$\phi(t) = (1 - 2it)^{-\frac{\nu}{2}}$